



Removal of Random Noise in Seismic Data by Time-varying Window-length Time-frequency Peak Filtering

Pengjun YU, Yue LI, Hongbo LIN, and Ning WU

Department of Information Engineering, Jilin University, Changchun, Jilin, China
e-mails: yupengjun1015@gmail.com, liyue@jlu.edu.cn (corresponding author),
hblin@jlu.edu.cn, Ning1337@gmail.com

Abstract

Time-frequency peak filtering (TFPF) is an effective tool for the removal of random noise and can be used to process seismic data with a low signal-to-noise ratio. A crucial aspect of this algorithm is the choice of window length (WL) of the time-frequency distribution. Whereas a fixed WL cannot simultaneously preserve signal and attenuate noise, time-varying WLs can achieve this goal. We propose a new method, L-DVV (delay vector variance), which successfully processes non-stationary signals by using the surrogate to measure the non-linearity of a time series. This method is sensitive to random noise and can accurately recover seismic signal masked by noise. Since the linearity criterion also meets the unbiased estimation criterion of the TFPF algorithm, the L-DVV method can be used for time-varying WL TFPF processing. Analysis of synthetic and real seismic data shows that the time-varying WL TFPF algorithm is effective at removing noise and recovering seismic signal.

Key words: TFPF, time-varying WL, L-DVV method, non-linearity, seismic random noise attenuation.

1. INTRODUCTION

Seismic data can be affected by both coherent and random noise (Groos and Ritter 2009). Coherent noise can be effectively removed during field processing by the wavelet method, singular value decomposition, and the recently proposed co-core trace transform method (see Lu 2006, Wu *et al.* 2011, for a review). Random noise can complicate both seismic prospecting and earthquake seismology. Many methods have been proposed for its removal including Wiener filtering, which performs well in removing some incoherent noise, and spatial predictive filtering in the f-k domain, which is effective at seismic noise suppression (Leite *et al.* 2008). However, these methods assume stationary signal conditions and cannot be applied in situations with non-stationary signals. Time-frequency peak filtering (TFPF) is a time-frequency based algorithm that first encodes the noisy time series as the instantaneous frequency (IF) of a frequency-modulated analytic signal and then performs IF estimation to recover the desired signal. The TFPF algorithm can estimate non-stationary signals without prior knowledge and can suppress random noise even in data with a low signal-to-noise ratio (SNR) (Boashash and Mesbah 2004). However, the use of a fixed window length (WL) in TFPF restricts both the preservation of the desired signal and the suppression of the random noise (see Gibbons *et al.* 2008, Liu *et al.* 2013, for a review). We therefore propose a new delay vector variance method based on the straight line sequence (L-DVV) to carry out TFPF with a time-varying WL.

To extract a seismic signal from random noise, the difference in non-linearity between the signal and the noise must be determined (Gautama *et al.* 2004a,b). The delay vector variance (DVV) method analyzes the non-linearity of a time series by comparing the original time series with a linearized “surrogate” version on a scatter diagram (Schreiber and Schmitz 1996). This method is superior to the AR model and the Volterra series for non-stationary signal processing (Galka and Ozaki 2001). However, the surrogate is commonly constructed by the iterative amplitude adjusted Fourier transformation (iAAFT) method, which can confuse the seismic signal and the random noise. The iAAFT of the surrogate has the same probability density distribution and roughly the same power spectrum as the original time series, while the noise (such as white Gaussian noise) has a roughly even power spectrum (Gautama *et al.* 2004b). The proposed delay vector variance method based on the straight line sequence (L-DVV) is essentially an extension of the DVV method (Yu *et al.* 2015). The straight line sequence is selected as the surrogate time series, meaning that this method is extremely sensitive to random noise and is able to accurately distinguish between the

seismic signal and noise, providing a basis for time-varying WL TFPF processing.

Assuming that the seismic signal is symmetric, the L-DVV method can determine the exact location of the wave crest, the remainder of the desired signal, and the random noise. The TFPF algorithm can therefore be applied with a time-varying WL such that reflection events can be recovered with a shorter WL and noise can be suppressed with a longer WL. The resulting time-varying WL TFPF is superior to conventional TFPF in both signal preservation and noise attenuation for synthetic and real seismic data.

In this letter, we introduce the TFPF algorithm in Section 2 and describe the time-varying WL TFPF algorithm based on the L-DVV method in Section 3. In Section 4 we apply the method to a synthetic record and a real shot from a crustal reflection experiment in Northeast China. Finally, we present our conclusions in Section 5.

2. TFPF ALGORITHM

The TFPF algorithm filters noisy signal by two procedures that seek to recover the desired signal using the instantaneous frequency (IF) in the frequency-modulated (FM) analytic signal (Boashash and Mesbah 2004).

In the first step, the analytic signal $z_s(t)$ is encoded by the IF of the noisy seismic signal $s(t)$ as follows:

$$z_s(t) = e^{j2\pi\mu \int_0^t s(\varphi) d\varphi}, \quad (1)$$

where μ is the scaling parameter analogous to the FM modulation index and $s(t)$ is the noisy seismic signal that could be described as a band-limited, non-stationary deterministic signal, given by

$$s(t) = x(t) + n(t) = \sum_{k=1}^q x_k(t) + n(t), \quad (2)$$

where $x_k(t)$ denotes the desired components of the signal and $n(t)$ is random noise.

In the second step, the filtered signal $\hat{x}(t)$ is generated by estimating the peak of the time-frequency distribution (TFD) of $z_s(t)$, which is expressed as follows:

$$\hat{x}(t) = f_{zs}(t) = \frac{\arg \max_f [W_{zs}(t, f)]}{\mu}, \quad (3)$$

where $W_{zs}(t, f)$ denotes the Wigner-Ville distribution (WVD) of $z_s(t)$ and is given by

$$W_{zs}(t, f) = \int_{-\infty}^{+\infty} z_s\left(t + \frac{\tau}{2}\right) z_s^*\left(t - \frac{\tau}{2}\right) e^{-j2\pi f \tau} d\tau. \quad (4)$$

Seismic random noise could be effectively attenuated by these two aforementioned procedures. For special cases in which the signal is linear in time and embedded in stationary white Gaussian noise, TFPF could give an absolutely unbiased estimation, and the bias $B(t)$ is written as (Wu *et al.* 2011)

$$B(t) = \arg \max_f \left[W_{zx}(t, f) \cdot \frac{4\pi^2 k_{n2} \mu^2}{(2\pi^2 k_{n2} \mu^2)^2 + (2\pi f - 2\pi \mu x(t))^2} \right] \cdot \frac{1}{\mu} - x(t) , \quad (5)$$

where $W_{zx}(t, f)$ denotes the WVD of $z_x(t)$ as Eq. 4 and k_{n2} is the second cumulate of the noise $n(t)$. In a particular case in which the desired signal is linear in time and could be described as $B(t) = \alpha t + C$ (where α and C are constants), $W_{zx}(t, f)$ and $B(t)$ could be simplified as follows:

$$W_{zx}(t, f) = \delta(f - \mu x(t)) . \quad (6)$$

$$B(t) = 0 . \quad (7)$$

Both Eqs. 5 and 7 are derived in the case of white Gaussian noise. Moreover, the whole analysis could still apply to other types of noise as long as the desired signal is linear in time, where the bias $B(t)$ in Eq. 7 is independent of k_{n2} . On the other hand, TFPF would result in a deterministic bias for common cases in which the signal is estimated as a finite-order polynomial or embedded in other types of noise.

Therefore, the pseudo-WVD (PW), which is a version of the windowed WVD, is calculated to realize the local linearity:

$$PW_{zs}(t, f) = \int_{-\infty}^{+\infty} h(\tau) z_s \left(t + \frac{\tau}{2} \right) z_s^* \left(t - \frac{\tau}{2} \right) e^{-j2\pi f \tau} d\tau , \quad (8)$$

where $h(\tau)$ is a rectangular window designed to modulate the local linearity. The optimal WL can be expressed as a function of the dominant frequency f_d and the sampling frequency f_s in the case of 5% deviation, as follows (Lin *et al.* 2008):

$$WL \leq \frac{0.384 f_s}{f_d} . \quad (9)$$

Filtering with a fixed WL does not necessarily result in simultaneous signal preservation and noise attenuation. To address this limitation, the time-varying WL TFPF algorithm calculates optimal WLs that correspond to different levels of data non-linearity. In the following sections, we describe the determination of the non-linearity of time-series data using the L-DVV method.

3. TIME-VARYING WL TFPF ALGORITHM BASED ON THE L-DVV METHOD

The newly proposed L-DVV method is essentially an extension of the DVV method with the surrogate linearized version of the time series selected as the straight line sequence. This method is superior to AR model for processing of non-stationary data and is highly sensitive to random noise. The linearity criterion also meets the unbiased estimation criterion of TFPF, so this method can also be applied to time-varying WL processing. The time-varying-WL TFPF algorithm based on the L-DVV method is summarized as follows.

Step (1) For the optimal embedding dimension m and time lag τ , a time series $\{x_i\} (i = 1, 2, \dots, N)$ can generate the delay vectors (DVs):

$$\bar{x}(k) = [x_{k-m\tau}, x_{k-(m-1)\tau}, \dots, x_{k-\tau}]^T. \quad (10)$$

Here, $\bar{x}(k)$ is the k th delay vector in the reconstructed phase space, x_k is the corresponding target, m is selected from the TFPF WL with the best performance (generally $WL_0 = 9$), and τ is set to 1 to meet the time-domain continuity criterion. In the reconstructed phase space, the pairwise Euclidean distances between DVs can be computed as follows:

$$d = \|\bar{x}(i) - \bar{x}(j)\| \quad (i \neq j). \quad (11)$$

The mean ρ_d and the standard deviation σ_d of the distances are subsequently computed. A set $\Omega_k(r_d)$ is generated by grouping the DVs that are within a certain Euclidean distance r_d from $\bar{x}(k)$:

$$\Omega_k(r_d) = \{\bar{x}(i) \mid \|\bar{x}(k) - \bar{x}(i)\| \leq r_d\}, \quad (12)$$

where r_d is taken from the interval $[\max\{0, \rho_d - n_d\sigma_d\}, \rho_d + n_d\sigma_d]$ and n_d is the parameter that controls the span.

Step (2) For the set $\Omega_k(r_d)$ the variance of the corresponding targets σ_k^2 is computed and normalized by the variance of the time series σ_s^2 to yield the target variance $\sigma_{(rd)}^2$:

$$\sigma_{(rd)}^2 = (1/N\sigma_s^2) \sum_{k=1}^n \sigma_k^2. \quad (13)$$

where N is the number of $\Omega_k(r_d)$. The DVV plot $(\sigma_{(rd)}^2)$ is a function of the standardized distance $(r_d - \rho_d)/\sigma_d$.

Step (3) For the time series $\{x_i\} (i = 1, 2, \dots, N)$ the surrogate time series of the L-DVV method $\{s_i\} (i = 1, 2, \dots, N)$ is a straight line sequence of the same length:

$$s_i = i \cdot \max(\{x_i\}) / N, \quad 1 \leq i \leq N. \quad (14)$$

The target variance of the surrogate time series $\sigma_{(rd)}^{*2}$ is computed in the first two steps. The horizontal axis of the scatter diagram corresponds to the DVV plot of the surrogate time series and the vertical axis corresponds to that of the original time series. Quantification of the delay vector variance (QDV) is defined as follows (Yu *et al.* 2015):

$$\text{QDV} = S / d. \quad (15)$$

As shown in Fig. 1, d is the distance between A and B, which are the projections of the end points of the scatter curve onto the bisector line. S is the area contained by the scatter curve and the bisector line. Therefore, by incorporating QDV, the degree of non-linearity is quantified and the analysis is simplified.

Step (4) The non-linearity of the time point t is defined as the QDV in the time interval $[\max(0, t - (N - 1) / 2), \min(t + (N - 1) / 2, L)]$, such that:

$$\text{QDV}(t) = \text{QDV}(\max(0, t - (N - 1) / 2), \min(t + (N - 1) / 2, L)), \quad (16)$$

where L is the time length and N reflects the time length of the interval. If the interval length is longer than the length of the desired signal, the non-linearity determination will be influenced by the adjacent $(N - 1) / 2$ points.

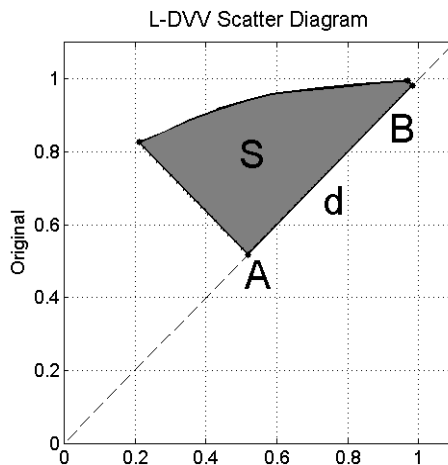


Fig. 1. Schematic representation of QDV on a scatter diagram.

To overcome this problem, each time point in the interval ($T(t)$) is marked one more time than the former, if the QDV of the time point t is less than the threshold value θ :

$$\begin{cases} T(t) = T(t) + 1, \text{QDV}(t) > \theta \\ T(t) = T(t), \text{QDV}(t) \leq \theta \end{cases}, \quad (17)$$

where $T(t)$ is the marked time of point t . Points with marked times more than $0.99N$ are filtered with a short WL (WL_{\min}), those with marked times in the range $[0.75N, 0.99N]$ are filtered with a moderate WL (WL_0), and the remaining points are filtered with a long WL (WL_{\max}).

$$\begin{cases} WL(t) = WL_{\min}, T(t) \in [0.99N, N] \\ WL(t) = WL_0, T(t) \in [0.75N, 0.99N] \\ WL(t) = WL_{\max}, T(t) \in [0, 0.75N] \end{cases}. \quad (18)$$

Therefore, this algorithm performs a trade-off between signal preservation and noise attenuation in time-varying WL processing.

4. APPLICATION TO SEISMIC RECORDS

4.1 Synthetic seismic records

To investigate the ability of the algorithm to remove noise from seismic data, we selected a 40-channel synthetic seismic record that contains two reflection events with dominant frequencies of 40 and 25 Hz and velocities of 2200 and 2800 m/s, respectively, which are embedded in white Gaussian noise with a SNR of -7 dB (Fig. 2a and b). For the time-varying WL TFPF, the short, moderate, and long WLs are set to 5, 9, and 31 samples, respectively. The TFPF without WL time-variation is set to 9 samples.

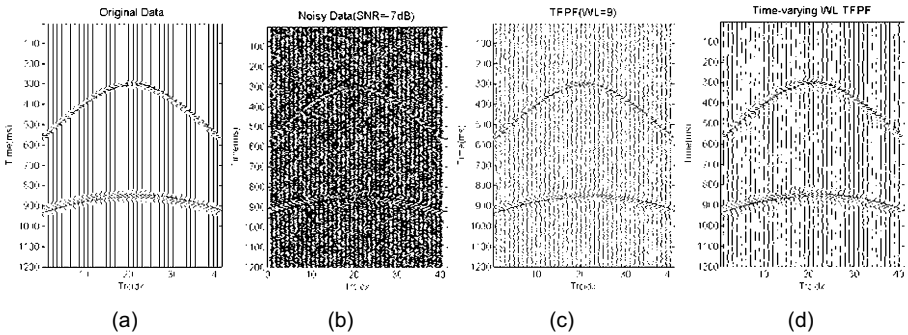


Fig. 2. Analysis of synthetic seismic data by TFPF and time-varying WL TFPF: (a) the original 40-trace record, (b) the seismic record with added Gaussian noise, (c) result of filtering with standard TFPF, and (d) result of filtering with the time-varying WL TFPF.

Figure 2c and d shows the results of filtering the synthetic seismic record. Both algorithms suppress most of the random noise and allow the two events to be recognized, but the time-varying WL TFPF produces a cleaner background with more continuous events. The SNRs of the time-varying WL and standard TFPF results are 0.9586 and 11.7087 dB, respectively, which demonstrates the superior noise suppression performance of the proposed method. More detailed comparisons were carried out on the tenth trace of the record, and Fig. 3 shows the QDV of each point, the marked times from the QDVs, the results of extraction by the L-DVVF method, and the filtered results. A comparison between wave crests for the original, noisy, standard TFPF and time-varying WL TFPF data (Fig. 4a and b) shows that the standard and time-varying WL TFPF algorithms reproduce the wave crest amplitude of the 40 Hz reflection event to 71.82 and 92.39%, respectively, and reproduce the wave crest amplitude of the 25 Hz reflection event to 78.54 and 88.67%, respectively. The time-varying WL TFPF algorithm produces better noise suppression in both the time and frequency domains than the standard TFPF algorithm, especially for the 50 Hz frequency component (Fig. 4c and d).

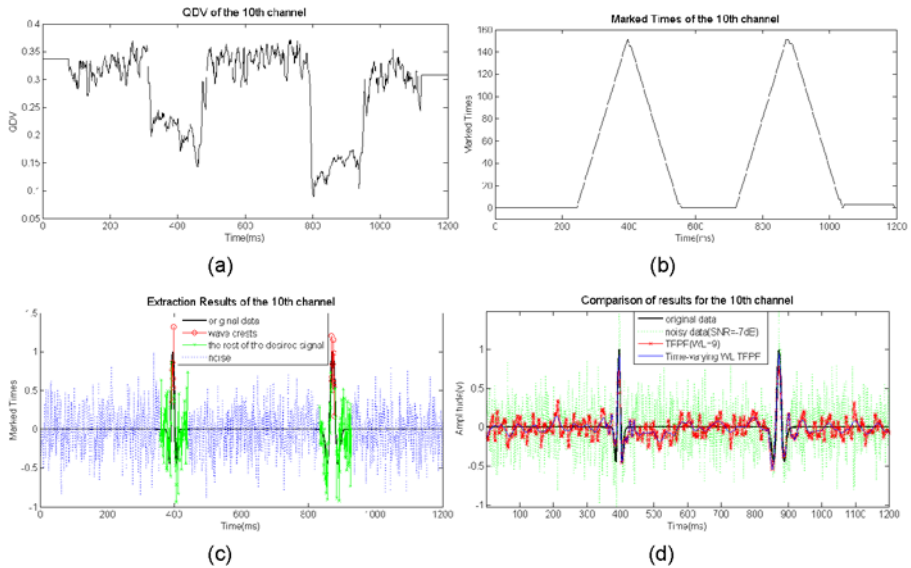


Fig. 3. Detailed comparison of the results of filtering the tenth trace of the synthetic data by TFPF and time-varying WL TFPF: (a) QDV of every point, (b) marked times according to the QDVs, (c) extraction result by the L-DVVF method, and (d) comparison of the results of both algorithms in the time domain.

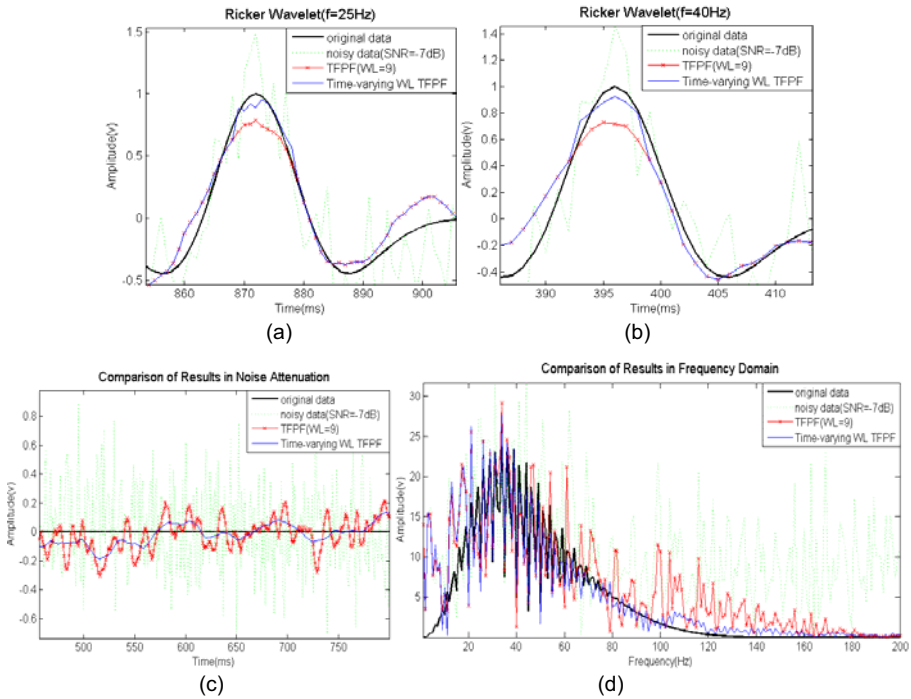


Fig. 4. Comparison between the results of filtering the tenth trace of the synthetic data by TFPF and time-varying WL TFPF: (a) amplitude preservation for the 40-Hz reflection event, (b) amplitude preservation for the 25-Hz reflection event, (c) noise suppression in the time domain, and (d) noise suppression in the frequency domain.

4.2 Field seismic data

Further comparisons between the algorithms were carried out using field seismic data from China comprising a 168 trace common-shot-point record with a record length of 6 s, sampled frequency of 1000 Hz, and geophone interval of 30 m (Fig. 5a). The WLs for both algorithms were set to the same values as the synthetic processing. The TFPF processing has removed much of the noise (Fig. 5b) but the events are still partly obscured by the remaining noise. In contrast, the time-varying WL TFPF results in better noise removal and more continuous recovered events (Fig. 5c). This is highlighted in the enlarged sections of the records and especially in the sections of the traces outlined by boxes in Fig. 5d-f, which show that the time-varying WL processing produces stronger wavelets with better continuity.

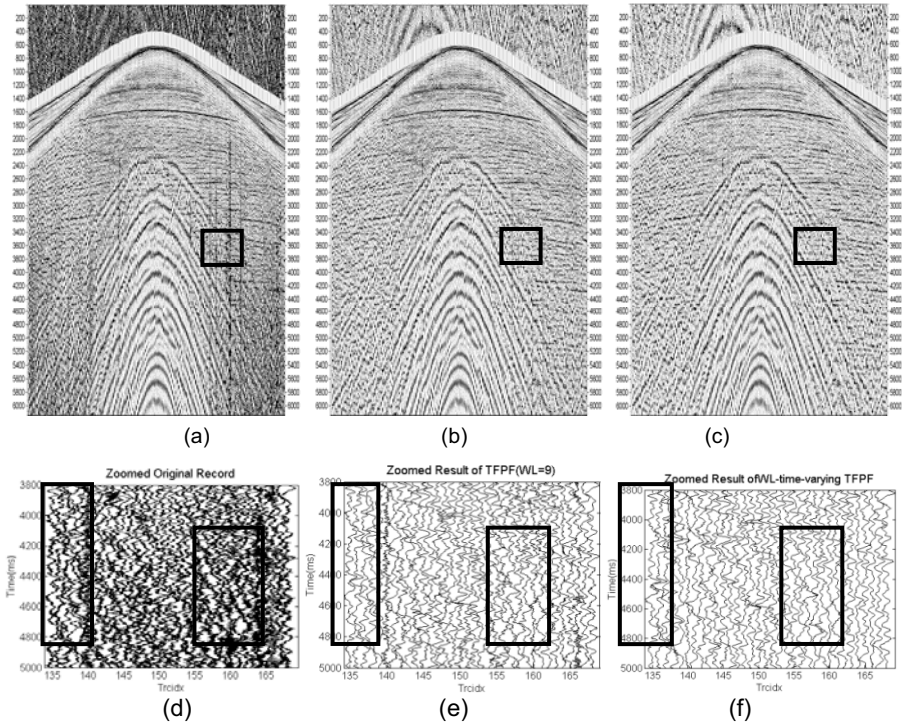


Fig. 5. Results of filtering of field data: (a) original noisy record, (b) result of filtering with the TFPF algorithm, (c) result of filtering with the time-varying WL TFPF algorithm, (d) enlargement of part of the original noisy record, (e) enlargement of part of the TFPF result, and (f) enlargement of part of the time-varying WL TFPF result.

5. CONCLUSIONS

The new L-DVV method produces non-stationary signal processing results that are superior to AR model by selecting the straight line sequence as the surrogate time series. It is sensitive to random noise and is able to extract a desired seismic signal from the noise. This method is also able to overcome the limitations of fixed WL TFPF by utilizing a time-varying WL since the linearity criterion also meets the unbiased estimation criterion of the TFPF algorithm. The resulting time-varying WL TFPF processing satisfactorily preserves the signal and attenuates the noise for both synthetic and real seismic data.

Acknowledgments. This research was supported by the National Natural Science Foundation of China (Grant 41130421 and Grant 41274118).

References

- Boashash, B., and M. Mesbah (2004), Signal enhancement by time-frequency peak filtering, *IEEE Trans. Signal Process.* **52**, 4, 929-937, DOI: 10.1109/TSP.2004.823510.
- Galka, A., and T. Ozaki (2001), Testing for nonlinearity in high-dimensional time series from continuous dynamics, *Physica D* **158**, 1-4, 32-44, DOI: 10.1016/S0167-2789(01)00318-9.
- Gautama, T., D.P. Mandic, and M. Van Hulle (2004a), The delay vector variance method for detecting determinism and nonlinearity in time series, *Physica D* **190**, 3-4, 167-176, DOI: 10.1016/j.physd.2003.11.001.
- Gautama, T., D.P. Mandic, and M.M. Van Hulle (2004b), A novel method for determining the nature of time series, *IEEE Trans. Biomedic. Eng.* **51**, 5, 728-736, DOI: 10.1109/TBME.2004.824122.
- Gibbons, S.J., F. Ringdal, and T. Kvaerna (2008), Detection and characterization of seismic phases using continuous spectral estimation on incoherent and partially coherent arrays, *Geophys. J. Int.* **172**, 1, 405-421, DOI: 10.1111/j.1365-246X.2007.03650.x.
- Groos, J.C., and J.R.R. Ritter (2009), Time domain classification and quantification of seismic noise in an urban environment, *Geophys. J. Int.* **179**, 2, 1213-1231, DOI:10.1111/j.1365-246X.2009.04343.x.
- Leite, F.E.A., R. Montagne, G. Corso, G.L. Vasconcelos, and L.S. Lucena (2008), Optimal wavelet filter for suppression of coherent noise with an application to seismic data, *Physica A* **387**, 7, 1439-1445, DOI: 10.1016/j.physa.2007.08.062.
- Lin, H.B., Y. Li, and B.J. Yang (2008), Varying-window-length time-frequency peak filtering and its application to seismic data. **In:** *Int. Conf. on Computational Intelligence and Security CIS '08, 13-17 December 2008*, 429-432, DOI: 10.1109/CIS.2008.160.
- Liu, Y.P., Y. Li, P.F. Nie, and Q. Zeng (2013), Spatiotemporal time-frequency peak filtering method for seismic random noise reduction, *IEEE Geosci. Remote Sens. Lett.* **10**, 4, 756-760, DOI: 10.1109/LGRS.2012.2221676.
- Lu, W.K. (2006), Adaptive noise attenuation of seismic images based on singular value decomposition and texture direction detection, *J. Geophys. Eng.* **3**, 1, 28-34, DOI: 10.1088/1742-2132/3/1/004.
- Schreiber, T., and A. Schmitz (1996), Improved surrogate data for nonlinearity tests, *Phys. Rev. Lett.* **77**, 4, 635-638, DOI: 10.1103/PhysRevLett.77.635.

- Wu, N., Y. Li, and B. Yang (2011), Surface wave attenuation of seismic records with the co-core trace transform filter, *Geophysics* **76**, 6, V115-V128, DOI: 10.1190/geo2010-0010.1.
- Yu, P.J., Y. Li, Y.L. Cheng, and H.B. Lin (2015), The L-DVV method for the seismic signal extraction, *J. Appl. Geophys.* **177**, 60-66, DOI: 10.1016/j.jappgeo.2015.03.016.

Received 5 January 2016

Received in revised form 12 June 2016

Accepted 15 June 2016